

Optimal Choice of Weights for Sparse Recovery With Prior Information

Axel Flinth*

May 25, 2016

Abstract

Compressed sensing deals with the recovery of sparse signals from linear measurements. Without any additional information, it is possible to recover an s -sparse signal using $m \gtrsim s \log(d/s)$ measurements in a robust and stable way. Some applications provide additional information, such as on the location of the support of the signal. Using this information, it is conceivable the threshold amount of measurements can be lowered. A proposed algorithm for this task is *weighted ℓ_1 -minimization*. Put shortly, one modifies standard ℓ_1 -minimization by assigning different weights to different parts of the index set $[1, \dots, d]$. The task of choosing the weights is however non-trivial.

This paper provides a complete answer to the question of an optimal choice of the weights. In fact, it is shown that it is possible to directly calculate unique weights that are optimal in the sense that the threshold amount of measurements needed for exact recovery is minimized. The proof uses recent results about the connection between convex geometry and compressed sensing-type algorithms.

Keywords: Compressed sensing, Gaussian matrices, weighted basis pursuit, convex geometry.

1 Introduction

At the heart of the theory of compressed sensing is the paradigm that it is possible to recover a sparse vector $x_0 \in \mathbb{R}^d$ using few linear measurements. A widely used method to perform this reconstruction is Basis Pursuit, i.e. to solve the minimization problem

$$\min \|x\|_1 \quad \text{subject to } Ax = b. \quad (P_1)$$

A fundamental result is that $m \gtrsim s \log d/s$ Gaussian random measurements are sufficient for Basis Pursuit to succeed at recovering an s -sparse vector with overwhelmingly high probability[2].

The main problem about recovering a signal in this setting is in fact to decide where its support $\text{supp } x_0$ is situated. Indeed, if we knew the exact position of $\text{supp } x_0$, we would be able to recover x_0 by simply solving a set of linear equations which have more equations than unknowns, provided $m \geq s$. It is thus plausible that if we are given prior information about the location of $\text{supp } x_0$, it should be possible to use this information to lower the threshold amount of measurements needed to secure recovery. To be precise, let us assume that we know that a fraction α of the indices in some set S are in the support of x_0 . This can be interpreted as the probability for an $i \in S$ to be an element of $\text{supp } x_0$. We will subsequently denote $T = \text{supp } x_0$, for the sake of brevity. Intuitively, it seems that if α is high while $\rho := |S|/d$ is not too large, we should be able to find the solution using less equations than if we did not have the initial guess.

*This work was supported by the Deutsche Forschungsgemeinschaft, DFG under Grant KU 1446/18-1. The author is with the Institut für Mathematik, Technische Universität Berlin, Germany. (e-mail: flinth@math.tu-berlin.de).

This exact situation was investigated in the paper [3]. Their proposed solution is to use a weighted ℓ_1 -approach, namely to solve the minimization problem

$$\min \|x\|_{1,w} \text{ subject to } Ax = b, \quad (P_1^w)$$

where $\|\cdot\|_{1,w}$ denotes the weighted ℓ_1 -norm, given by

$$\|x\|_{1,w} = \sum_{i=1}^d w_i |x(i)|$$

($\|\cdot\|_{1,w}$ should of course not be confused with the weak ℓ_1 -norm.) We can likewise define the reweighted ℓ_p -norms for every $p \in [1, \infty]$;

$$\|x\|_{p,w} = \left(\sum_{i=1}^d w_i |x(i)|^p \right)^{\frac{1}{p}}, \quad p < \infty \text{ and } \|x\|_{\infty,w} = \sup_i w_i |x(i)|.$$

Choosing the weights w_i is a subtle process – but the general idea is that the weights on the indices S should be lower. This will have the effect that high values $x(i)$ on S are not penalized, which is desirable if $T \approx S$. Many papers have been concerned with the choice $w = 1_{S^c}$ (the setting corresponding to this choice is sometimes called *modified Compressed Sensing* [10]), but the authors of [3] suggest that one should give oneself the freedom to choose w as $\omega 1_S + 1_{S^c}$ for some $\omega \in [0, 1]$. Here, as well as in the following, 1_M refers to the indicator function of a set M , i.e.

$$1_M(i) = \begin{cases} 1 & \text{if } i \in M \\ 0 & \text{else.} \end{cases}$$

Tedious arguments by the authors of [3] showed that provided that the guess is good ($\alpha > .5$), the required bound on the RIP-constant for the matrix A needed for robust recovery can be softened, and the stability constants get smaller, if one chooses $\omega = 0$. The formulas actually suggest that one either should use $\omega = 0$ or $\omega = 1$, and never something in between. Numerical experiments do not harmonize with this rule of thumb: in fact, if the guess is bad ($\alpha < .5$), one does significantly better choosing ω between 0 and 1 than choosing one of the extreme values. The authors left it as an open problem to analyze how to choose the weight in an optimal way.

1.1 Previous results on the optimal choice of weights.

It is of course not entirely clear what “optimal” means in this context. One way of defining it is to say that a weight is optimal if the minimal amount of measurements needed in order for (P_1^w) with $b = Ax_0$ to recover x_0 is as small as possible. A popular model assumption is that A is chosen according to some probability distribution. Because of its universality as a limit distribution, the Gaussian probability distribution is probably the most canonical one. In this setting, we search for the minimal amount of measurements required for P_1^w to succeed with high probability (the exact meaning of “high” varies from application to application).

The first result concerning choosing ω optimally in this sense was probably Theorem 4.3.3 of the PhD-thesis [11] (the result was also presented in [6]). Based on a very sophisticated argument relying on calculating internal and external angles of certain polytopes, the authors implicitly provide a threshold $m_0(\omega)$ so that if one uses $m \geq m_0(\omega)$ Gaussian measurements, (P_1^w) will succeed with high probability. One can then minimize $m_0(\omega)$ with respect to the weight ω . Since the formulas are very complicated, this is however a tough task.

A result which is simpler to grasp is given in [8]. The strategy is basically the same. One computes a threshold depending on the weight ω , using the famous “escape through a mesh lemma” [4] and then

minimizes the threshold with respect to ω . Although the paper provides powerful and beautiful results on how the optimal threshold is depending on α and ρ , it does not directly provide any method of actually calculating the optimal weights. Furthermore, as the authors point out, the result is only about an upper bound on how many measurements one needs.

A different analysis conducted in [7]. The authors define a *weighted Null Space Property* and prove several interesting result about that property. In particular, they calculate upper bounds on thresholds on the number of measurements needed to secure that Gaussian matrices have that property. The main technical tool is again Gordon's escape through a mesh lemma. As a concrete weighting strategy, it is proposed to set the weight $\omega = 1 - \alpha$, since that weight minimizes their bound on the thresholds. One should note that the weighted Null Space property is uniform in nature in the sense that it is equivalent to that all signals supported on a set S will be recovered from its linear measurements through weighted ℓ_1 -minimization as long as the quality of the support estimate is good enough - its exact position does not matter.

1.2 Contributions of this paper

In this paper, we will consider the full model with k subsets S_i of the complete index set $[1, \dots, d]$, each one with a corresponding probability α_i . Correspondingly, we will allow for k weights ω_i , one on each S_i . We will use the short-hand notation $\omega = (\omega_1, \dots, \omega_k)$. The strategy of finding the optimal weights is very similar to the ones of the papers mentioned above – we will calculate a threshold m_0 depending on ω and then prove that $m_0(\omega)$ is minimized for an optimal set of weights. To be precise, we will in fact not minimize m_0 itself, but only an upper and lower bound on m_0 . These bounds are however simultaneously minimized for the same ω^* and also lie very close to each other. We will also investigate how the mentioned bounds on m_0 depend on the properties of the sets S_i , $i = 1, \dots, k$. In particular, we will find the perhaps somewhat remarkable fact that one needs in principle as many measurements to recover an $(\alpha_1 |S_1| + \dots + \alpha_k |S_k|)$ -sparse signal given the prior knowledge of the probabilities $\alpha = (\alpha_1, \dots, \alpha_k)$ and the sets (S_1, \dots, S_k) as recovering k signals x_i separately, where for each i , x_i is $\alpha_i |S_i|$ -sparse and known to be supported on the set S_i . This was already proven for upper bounds on the measurements needed in [8], but here, we provide also a statement about lower bounds.

In order to do this, we will use a powerful and recent result from [1], namely that the recovery probability of a convex program undergoes a *phase transition* as the number of measurements surpass the *statistical dimension* $\delta(C)$ of a certain convex cone. We will provide a way of calculating a set of weights ω^* which simultaneously minimizes a tight *lower and upper* bound of $\delta(C)$, and hence to some extent settle the problem regarding the optimal choice of weights when recovering a sparse vector x_0 with prior information. Additionally, we will even provide a simple analytical formula for ω^* .

The fact that we are using the new notion of statistical dimensions is the main technical difference from the prior work mentioned in the previous section, where the escape through a lemma was the main tool in most cases.

The notation in the paper will be fairly standard. There are however a few things that should be pointed out. For a vector $w \in \mathbb{R}^d$, $\text{diag}(w)$ is the diagonal matrix in $\mathbb{R}^{d,d}$ whose diagonal is w . $\text{pos}(x)$ is the positive part of a real number, i.e. $\text{pos}(x) = x$ if $x \geq 0$, and otherwise $\text{pos}(x) = 0$. For $x \in \mathbb{R}^d$, $\text{sgn } x$ denotes the vector consisting of the signs of the entries of x , with the convention $\text{sgn } 0 = 0$. Finally, $|S|$ will have different meanings depending on its argument. If S is a real number, $|S|$ denotes its absolute value. For vectors, $|S|$ is the vector formed by taking the absolute value pointwise. If S is a set, $|S|$ denotes the number of elements of S .

2 Statistical dimensions and the measurement threshold

Let us begin by fixing the model situation and notation. $x_0 \in \mathbb{R}^d$ is an s -sparse vector, meaning that only s of the entries $x_0(i)$, $i = 1, \dots, d$ are unequal to zero. Additional to the prior information that x_0 is sparse,

we are given some partial information where the support T of x_0 is situated, expressed by the partition $\mathcal{S} = (S_i)_{i=1}^k$ of $[1, \dots, d]$. With each set S_i we associate the two parameters α_i and ρ_i

$$\rho_i = \frac{|S_i|}{d} \text{ and } \alpha_i = \frac{|T \cap S_i|}{|S_i|},$$

reflecting its size and the probability for a $j \in S_i$ to be an element in T , $i = 1, \dots, k$, respectively. To each S_i , we will assign a weight $\omega_i \geq 0$ and always choose the weights according to

$$w = \sum_{i=1}^k \omega_i 1_{S_i}.$$

We will without loss of generality assume that $\alpha_i > 0$ for every i (if $\alpha_i = 0$ for some i , the corresponding set of indices S_i can be completely ignored in the recovery process.)

Now we define the object which we will search for in the rest of this section: the threshold amounts of measurements needed for the program (P_1^w) to be successful.

Definition 1. Let $A \in \mathbb{R}^{m,d}$ be a Gaussian matrix. Then $\mu_{s,d}$ denotes the normalized threshold amount of measurement needed for (P_1) with $Ax_0 = b$ to succeed at recovering an s -sparse vector x_0 , i.e., (P_1) will succeed with high probability if $m \geq d(\mu_{d,s} + o(1))$, and will fail with high probability if $m \leq d(\mu_{d,s} - o(1))$.

In the exact same way, we define for a given support T and partition $\mathcal{S} = (S_i)_{i=1}^k$ the number $m_{\mathcal{S},T,w}$ as the threshold amount of measurements for (P_1^w) with $Ax_0 = b$ to succeed at recovering a vector x_0 supported on T . Finally, $m_{\mathcal{S},T}$ is the value obtained by choosing the weights w optimally, i.e. $m_{\mathcal{S},T} = \inf_{w \geq 0} m_{\mathcal{S},T,w}$.

Note that we are only concerned with exact recovery of exactly sparse signals, and, in particular, leave out cases where noise is involved and the signals are only approximately sparse. We believe that the requirement of stability and robustness will not significantly increase the measurement threshold, but also think that a thorough analysis of these questions are beyond the scope of this paper. Therefore they are left as possible lines of future research.

The authors of [8] provided upper bounds $\hat{\eta}_{s,d}$ and $\hat{\eta}_{\mathcal{S},T}$ of $\mu_{s,d}$ and $m_{\mathcal{S},T}$, respectively. (Note that our notation is a bit different to theirs: what we call $\hat{\eta}_{\mathcal{S},T}$ is denoted $\hat{\eta}(\alpha, \beta)$ by them, where α and β denotes $|S_1 \cap T|$ and $|S_2 \cap T|$, respectively.) They then proceed to prove the following very beautiful result.

Theorem 2. [8, Th. 3.1] We have

$$\hat{\eta}_{(S_1, S_2), T} = \rho_1 \hat{\eta}_{\alpha_1 |S_1|, |S_1|} + \rho_2 \hat{\eta}_{\alpha_2 |S_2|, |S_2|}$$

Theorem 2 in fact states that in order to find an $\alpha_1 \rho_1 d + \alpha_2 \rho_2 d$ -sparse vector using the prior information, we need about as many measurements as the sum of the measurements needed to separately recover two $\alpha_i \rho_i d$ -sparse vectors known to be supported on S_i , respectively. The authors of [8] conjectured that the theorem can be generalized to more than $k = 2$ sets S_i . We will do this at the end of the paper, with the addition that the upper bound we provide on $m_{\mathcal{S},T}$ in fact also induces a lower bound.

$\mu_{s,d}$ and $m_{\mathcal{S},T,w}$ can be calculated using a few results from [1]. Before stating the results we will use, let us define the statistical dimension of a cone C .

Definition 3. [1] Let $C \subseteq \mathbb{R}^d$ be a closed convex cone and g be a Gaussian vector. Let furthermore Π_C denote the metric projection onto C , i.e.

$$\Pi_C(x) = \operatorname{argmin}_{c \in C} \|c - x\|_2.$$

The statistical dimension $\delta(C)$ of C is then defined as

$$\delta(C) = \mathbb{E} \left(\|\Pi_C g\|_2^2 \right).$$

The statistical dimension of a general convex cone is the statistical dimension of its closure.

The main result of [1] shows that the statistical dimension of the descent cone $\mathcal{D}(f, x_0)$ of a convex function f at x_0 , i.e.,

$$\mathcal{D}(f, x_0) = \{v \in \mathbb{R}^d \mid \exists \tau > 0 : f(x_0 + \tau v) \leq f(x_0)\},$$

can be used to calculate the minimal amount of measurements needed for the program

$$\min f(x) \text{ subject to } Ax = b \quad (P_f)$$

with $b = Ax_0$ to succeed at recovering a vector $x_0 \in \mathbb{R}^d$. It is well known that (P_f) has x_0 as its solution if and only if $\mathcal{D}(f, x_0) \cap \ker A = \{0\}$. This leads to the following precise result:

Theorem 4. [1, Th. I] *Let $C \subseteq \mathbb{R}^d$ be a convex cone and $A \in \mathbb{R}^{m,d}$ a Gaussian matrix. Then if*

$$\delta(C) + \sqrt{8 \log(4/\eta)} \sqrt{d} \leq m,$$

we have $\mathbb{P}(C \cap \ker A \neq \{0\}) \leq \eta$. Likewise if

$$\delta(C) - \sqrt{8 \log(4/\eta)} \sqrt{d} \geq m,$$

then $\mathbb{P}(C \cap \ker A \neq \{0\}) \geq 1 - \eta$.

In particular, the probability that the solution of the problem (P_f) is equal to x_0 undergoes a phase transition as m surpasses $\delta(\mathcal{D}(f, x_0))$, and hence we have

$$\mu_{s,d} = \frac{\delta(\mathcal{D}(\|\cdot\|_1, x_0))}{d} \text{ and } m_{\mathcal{S},T,w} = \frac{\delta(\mathcal{D}(\|\cdot\|_{1,w}, x_0))}{d}.$$

The authors of [1] proceeded to calculate $\delta(\mathcal{D}(\|\cdot\|_1, x_0))$ as an exemplary application of a technique they presented. We will use the same technique to calculate $\delta(\mathcal{D}(\|\cdot\|_{1,w}, x_0))$, but before that, let us state the result regarding standard Basis Pursuit.

Theorem 5. [1, Prop. 4.5]. *Define the function J_σ through*

$$J_\sigma(\tau) = \sigma(1 + \tau^2) + (1 - \sigma)\varphi(\tau),$$

where $\varphi : [0, 1] \rightarrow [0, 1]$ is defined through

$$\varphi(t) = \sqrt{\frac{2}{\pi}} \int_t^\infty (x - t)^2 \exp(-x^2/2) dx. \quad (1)$$

The statistical dimension of the descent cone of the ℓ_1 -norm at an s -sparse $x_0 \in \mathbb{R}^d$ satisfies

$$\inf_{\tau > 0} J_\sigma(\tau) - \frac{2}{d} \sqrt{\frac{1}{\sigma}} \leq \frac{\delta(\mathcal{D}(\|\cdot\|_1, x_0))}{d} \leq \inf_{\tau > 0} J_\sigma(\tau),$$

where $\sigma = s/d$.

Now as promised, we will prove a similar result regarding the weighted Basis Pursuit $(P_{1,w})$. We will use the same technique as the authors of [1] and therefore start by calculating the expected length of the projection of a Gaussian vector onto a dilation of the subdifferential of the weighted ℓ_1 -norm. The *subdifferential* [9, p.214-15] of a convex function f at a point x_0 is the set of vectors p satisfying

$$f(x + v) \geq f(x_0) + \langle p, v \rangle \text{ for all } v \in \mathbb{R}^d.$$

For this, we will need a fact on subdifferentials of norms. The following result is folklore in this area, but let us include a proof for completeness.

Proposition 6. For any norm $\|\cdot\|$ on \mathbb{R}^d , denote by $\|\cdot\|_*$ its dual norm, i.e.

$$\|p\|_* = \sup_{\|v\|=1} \langle p, v \rangle.$$

Then for every $x_0 \neq 0$

$$\partial_{x_0} \|\cdot\| = \{p \mid \|p\|_* = 1, \langle p, x_0 \rangle = \|x_0\|\},$$

where $\langle p, x_0 \rangle$ denotes the canonical dual pairing $\mathbb{R}^d \times \mathbb{R}^d$, i.e., the Euclidean scalar product.

Proof. We aim to characterize the vectors $p \in \mathbb{R}^d$ which have the property

$$\|x_0 + v\| \geq \|x_0\| + \langle p, v \rangle \text{ for all } v \in \mathbb{R}^d \quad (2)$$

By taking the supremum in (2) over all v with $\|v\| = 1$, we immediately obtain

$$\sup_{\|v\|=1} \langle p, v \rangle \leq \sup_{\|v\|=1} \|x_0 + v\| - \|x_0\| \leq \sup_{\|v\|=1} \|v\| = 1,$$

i.e. $\|p\|_* \leq 1$. By further choosing $v = -x_0$, we get $\langle p, x_0 \rangle \geq \|x_0\|$, which proves $\|p\|_* \geq 1$, and therefore also $\|p\|_* = 1$. This further implies that $\langle p, x_0 \rangle = \|x_0\|$. If on the other hand $\|p\|_* = 1$ and $\langle p, x_0 \rangle = \|x_0\|$, we have

$$\|x_0\| + \langle p, v \rangle = \langle p, x_0 + v \rangle \leq \|x_0 + v\|,$$

and hence (2) is true, and we are done. \square

Now we can calculate $\mathbb{E} \left(\text{dist}^2(g, \tau \cdot \partial_{x_0} \|\cdot\|_{1,w}) \right)$ for $x_0 \in \mathbb{R}^d$ and a Gaussian vector g .

Lemma 7. Let $x_0 \in \mathbb{R}^d$ be arbitrary and g a Gaussian vector. We have

$$\mathbb{E} \left(d^{-1} \text{dist}^2(g, \tau \cdot \partial_{x_0} \|\cdot\|_{1,w}) \right) = \sigma + \sum_{i=1}^k \rho_i (\alpha_i (\omega_i \tau)^2 + (1 - \alpha_i) \varphi(\omega_i \tau)),$$

where $\sigma = \|x_0\|_0 / d$, $w = \sum_{i=1}^k \omega_i 1_{S_i}$ and φ is defined through Equation (1).

Proof. Let us start by noticing that the dual norm of $\|\cdot\|_{1,w}$ is $\|\cdot\|_{\infty, w^{-1}}$. The Hölder inequality implies that for every $x, v \in \mathbb{R}^d$,

$$\langle x, v \rangle = \sum_{j=1}^d w_j^{-1} x_j w_j v_j \leq \|(w_i^{-1} x_i)_{i=1}^d\|_{\infty} \|(w_i v_i)_{i=1}^d\|_1 = \|x\|_{\infty, w^{-1}} \|v\|_{1,w},$$

i.e. $\|x\|_{1,w}^* \leq \|x\|_{\infty, w^{-1}}$. By choosing v equal to the vector for which $v(i) = \text{sgn } x(i)$ if i is the index where $w(i)^{-1} x(i)$ is maximal, and else $v(i) = 0$, we even see that equality holds.

This, together with the fact that under the assumption $\|p\|_{w^{-1}, \infty} = 1$, $\langle x_0, p \rangle = \|x_0\|_{1,w}$ if and only if $p|_{\text{supp } x_0} = D^w \text{sgn } x_0|_{\text{supp } x_0}$, implies that

$$\partial_{x_0} \|\cdot\|_{1,w} = \left\{ D^w \text{sgn } x_0 + v \mid \text{supp } v \cap \text{supp } x_0 = \emptyset, \forall j : |v_j| \leq w_j \right\},$$

where $D^w = \text{diag}(w) = \text{diag} \left(\sum_{i=1}^k \omega_i 1_{S_i} \right)$. Abbreviating $T = \text{supp } x_0$, we see that

$$\text{dist}^2(g, \tau \cdot \partial_{x_0} \|\cdot\|_{1,w}) = \sum_{j \in T} (g(j) - \text{sgn}(x_0(j)) w_j \tau)^2 + \sum_{j \notin T} \text{pos}(|g(j)| - w_j \tau)^2$$

(the part on T of a vector in $\tau \partial_{x_0} \|\cdot\|_{1,w}$ is fixed to $\tau D^w \text{sgn } x_0$, and the other part can elementwise not be larger than w .) Taking the expectation, we obtain

$$\begin{aligned} \mathbb{E}(\text{dist}^2(g, \tau \cdot \partial_{x_0} \|\cdot\|_{1,w})) &= \sum_{j \in T} (1 + w_j^2 \tau^2) + \sum_{j \notin T} \sqrt{\frac{2}{\pi}} \int_{w_j \tau}^{\infty} (x - w_j \tau)^2 \exp(-x^2/2) dx \\ &= |T| + \sum_{i=1}^k |S_i \cap T| \tau^2 \omega_i^2 + |S_i \cap T^c| \varphi(\omega_i \tau) \\ &= d \left(\sigma + \sum_{i=1}^k \rho_i (\alpha_i (w_i \tau)^2 + (1 - \alpha_i) \varphi(\omega_i \tau)) \right). \end{aligned}$$

□

For convinence, let us abbreviate

$$J_{\sigma, \alpha, \rho, \omega}(\tau) = \sigma + \sum_{i=1}^k \rho_i (\alpha_i (\omega_i \tau)^2 + (1 - \alpha_i) \varphi(\omega_i \tau)).$$

Also define

$$\tilde{m}_{S, T, w} := \inf_{\tau > 0} J_{\sigma, \alpha, \rho, \omega}(\tau), \quad (3)$$

where α and ρ are the parameters associated with S and T , and $w = \sum_{i=1}^k \omega_i 1_{S_i}$.

As the notation suggests, $\tilde{m}_{S, T, w}$ is not far away from $m_{S, T, w}$. We will prove this using Lemma 7 together with the following theorem from [1].

Theorem 8. [1, Th. 4.3]. *Let $\|\cdot\|$ be a norm on \mathbb{R}^d . The the statistical dimension of the descent cone of $\|\cdot\|$ at a point $x_0 \in \mathbb{R}^d$ satisfies*

$$0 \leq \inf_{\tau > 0} \mathbb{E}(\text{dist}^2(g, \tau \partial_{x_0} \|\cdot\|)) - \delta(\mathcal{D}(\|\cdot\|, x_0)) \leq \frac{2 \sup_{s \in \partial_{x_0} \|\cdot\|} \|s\|_2}{\left\| \frac{x_0}{\|x_0\|_2} \right\|}. \quad (4)$$

Proposition 9. *The statistical dimension of $\mathcal{D}(\|\cdot\|_{1,w}, x_0)$ satisfies*

$$\tilde{m}_{S, T, w} - \frac{2}{d} \sqrt{\frac{1}{\bar{\alpha}}} \leq \frac{\delta(\mathcal{D}(\|\cdot\|_{1,w}, x_0))}{d} \leq \tilde{m}_{S, T, w}, \quad (5)$$

where $\bar{\alpha} = \min_i \alpha_i$ and $\tilde{m}_{S, T, w}$ is defined through (3). Moreover, the following bound independent of α holds:

$$\tilde{m}_{S, T, w} - \frac{2}{\sqrt{d}} \leq \frac{\delta(\mathcal{D}(\|\cdot\|_{1,w}, x_0))}{d} \leq \tilde{m}_{S, T, w}. \quad (6)$$

Proof. We use Theorem (8). To control the error term, by using that $|S_i \cap T| = \alpha_i \rho_i d$ for each i , we obtain the estimate

$$\|x\|_{1,w} = \sum_{j \in T} w_j |x_j| \leq \|x\|_2 \sqrt{\sum_{j \in T} w_j^2} = \sqrt{d \sum_{i=1}^k \rho_i \alpha_i \omega_i^2} \|x\|_2$$

which is valid for any x supported on T . We even have equality if we choose $|x|$ parallel to $w|_{S_0}$. We can do this since, as pointed out in [1], there is no need to use Theorem (8) directly for x_0 . We only have to

secure that x has the same $\mathcal{D}(\|\cdot\|_{1,w}, x)$ and $\partial_x \|\cdot\|_{1,w}$ as x_0 , since $\inf_{\tau>0} J_{\sigma,\alpha,\rho,\omega}(\tau) - \delta(\mathcal{D}(\|\cdot\|_{1,w}, x_0))$ only depends on those entities. It is clear from the considerations in the proof of Lemma 7 that for this to be true, x only has to be supported on T and have the same sign pattern as x_0 . It is equally clear that there exist such vectors x so that $|x|$ is parallel to $w|_T$.

This proves that it is possible to choose the denominator of the error term in Theorem (8) equal to $\sqrt{d \sum_{i=1}^k \rho_i \alpha_i \omega_i^2}$. To bound the numerator, it suffices to notice that each vector $s \in \partial_{x_0} \|\cdot\|_{1,w}$ satisfies

$$\|s\|_2 = \sqrt{\sum_{j=1}^d s_j^2} \leq \sqrt{\sum_{j=1}^d w_j^2} = \sqrt{\sum_{i=1}^k \sum_{j \in S_i} \omega_i^2} = \sqrt{d \sum_{i=1}^k \rho_i \omega_i^2},$$

since $|s_j| \leq w_j$ for each j , and $|S_i| = \rho_i d$. To finish the proof of the tighter bound (5), we note that

$$\frac{\sqrt{d \sum_{i=1}^k \rho_i \omega_i^2}}{\sqrt{d \sum_{i=1}^k \rho_i \alpha_i \omega_i^2}} \leq \sqrt{\frac{d \sum_{i=1}^k \rho_i \omega_i^2}{d \bar{\alpha} \sum_{i=1}^k \rho_i \omega_i^2}} = \frac{1}{\sqrt{\bar{\alpha}}}.$$

Finally, the bound (6) easily follows from that statement, since

$$d\bar{\alpha} = \min_i d \frac{|S_i \cap T|}{|S_i|} \geq 1,$$

using the facts that $|S_i \cap T| \geq 1$ and $|S_i| \leq d$. □

3 Optimizing $m_{\mathcal{S},T,w}$.

Proposition 9 states that $m_{\mathcal{S},T,w}$ is bounded both below and above by the number $\tilde{m}_{\mathcal{S},T,w} = \inf_{\tau>0} J_{\sigma,\alpha,\rho,w}(\tau)$, up to an error which is independent of the weights w . In particular, if we denote $\tilde{m}_{\mathcal{S},T} = \inf_{w \geq 0} \tilde{m}_{\mathcal{S},T,w}$, we have

$$\tilde{m}_{\mathcal{S},T} - \frac{2}{\sqrt{d}} \leq m_{\mathcal{S},T} \leq \tilde{m}_{\mathcal{S},T}.$$

Hence, $\tilde{m}_{\mathcal{S},T}$ is equal to $m_{\mathcal{S},T}$ up to an asymptotically vanishing error term. In this section, we prove that $\tilde{m}_{\mathcal{S},T,w}$ has a unique minimum at some ω^* for every \mathcal{S} . As the proof will show, it can furthermore easily be found by minimizing a convex function in two variables. The proof uses several techniques exploited for partly other purposes in [1].

Theorem 10. *Assume that $|T| < d$ and let $x_0 \in \mathbb{R}^d$ be a vector supported on T . For each partition $\mathcal{S} = (S_i)_{i=1}^k$ of $[1, \dots, d]$, there exists weights $\omega^* \in \mathbb{R}_+^k$ so that*

$$\tilde{m}_{\mathcal{S},T,\omega} = \inf_{\tau>0} \mathbb{E} \left(\text{dist}^2(g, \tau \partial_{x_0} \|\cdot\|_{1,w}) \right)$$

is minimal, where $w = \sum_{i=1}^k \omega_i 1_{S_i}$. The weights ω^ are for instance uniquely determined by the additional constraint $\|\omega^*\|_\infty = 1$.*

Proof. First, we note that the chain rule proves that

$$\partial_{x_0} \|\cdot\|_{1,w} = D^w (\partial_{D^w x_0} \|\cdot\|_1) = D^w (\partial_{x_0} \|\cdot\|_1) =: D^w C,$$

where we used $D^{w*} = D^w$ and the fact that the subdifferential $\partial_x \|\cdot\|_1$ only depends in the sign pattern of x . Hence, since x_0 and $D^w x_0$ have the same sign pattern, $\partial_{D^w x_0} \|\cdot\|_1 = \partial_{x_0} \|\cdot\|_1$. If for $v \in \mathbb{R}_+^k$, we denote $\Delta^v = \text{diag}(\sum_{i=1}^k v_i 1_{S_i})$, then $\tau \Delta^v = \Delta^{\tau \omega}$ and hence

$$\inf_{\omega>0} \inf_{\tau>0} \mathbb{E} \left(\text{dist}^2(g, \tau \partial_{x_0} \|\cdot\|_{1,w}) \right) = \inf_{v>0} \mathbb{E} \left(\text{dist}^2(g, \Delta^v C) \right).$$

Therefore, if we prove that there exists a unique $v^* \neq 0$ so that $\mathbb{E}(\text{dist}^2(g, \Delta^v C))$ is minimal, then $\omega^* = v^* / \|v^*\|_\infty$ is the optimal weight we are looking for, since then

$$\inf_{\tau > 0} \mathbb{E}(\text{dist}^2(g, \tau \partial_{x_0} \|\cdot\|_{1, \omega^*})) = \inf_{\tau > 0} \mathbb{E}(\text{dist}^2(g, \Delta^{\tau \omega^*} C)) = \mathbb{E}(\text{dist}^2(g, \Delta^{v^*} C)) \leq \mathbb{E}(\text{dist}^2(g, \Delta^{\tau \omega} C))$$

for any $\tau > 0$, $\omega \geq 0$, from which the optimality of ω^* follows. ω^* of course denotes $\sum_{i=1}^k \omega_i^* 1_{S_i}$.

In order to prove that there exists a unique minimizer of $\phi(v) = \mathbb{E}(\text{dist}^2(g, \Delta^v C))$ on \mathbb{R}_+^k , it suffices to prove that ϕ is strictly convex, continuous and coercive (“large outside a ball”).

Strict convexity: Let $s_i \in \Delta^{v_i} C$, $i = 1, 2$, and $\theta \in [0, 1]$. Then (due to the structure of $C = \partial_{x_0} \|\cdot\|_1$), we know that $s_i = \Delta^{v_i}(\text{sgn } x_0 + u_i)$ for some u_i with $\text{supp } u_i \cap \text{supp } x_0 = \emptyset$ and $\|u_i\|_\infty \leq 1$. This implies that

$$\theta s_1 + (1 - \theta)s_2 = (\theta \Delta^{v_1} + (1 - \theta) \Delta^{v_2}) \text{sgn } x_0 + (\theta \Delta^{v_1} u_1 + (1 - \theta) \Delta^{v_2} u_2).$$

The right hand side of the previous equation is an element of $\Delta^{\theta v_1 + (1 - \theta)v_2} C$. This since $\theta \Delta^{v_1} + (1 - \theta) \Delta^{v_2} = \Delta^{\theta v_1 + (1 - \theta)v_2}$, the action of diagonal matrices does not change the support of a vector, and the chain of inequalities

$$\begin{aligned} |\langle \theta \Delta^{v_1} u_1 + (1 - \theta) \Delta^{v_2} u_2, e_i \rangle| &\leq \theta |\langle \Delta^{v_1} u_1, e_i \rangle| + (1 - \theta) |\langle \Delta^{v_2} u_2, e_i \rangle| \\ &= \theta |u_1(i)| v_1(i) + (1 - \theta) |u_2(i)| v_2(i) \leq \theta v_1(i) + (1 - \theta) v_2(i). \end{aligned}$$

From the latter we deduce that $\theta \Delta^{v_1} u_1 + (1 - \theta) \Delta^{v_2} u_2$ can be written as $\Delta^{\theta v_1 + (1 - \theta)v_2} u$, where $\text{supp } u \cap \text{supp } x_0 = \emptyset$ and $\|u\|_\infty \leq 1$. We have thus proven that for every triple $v_1, v_2 \in \mathbb{R}_+^k$, $\theta \in [0, 1]$,

$$\theta \Delta^{v_1} C + (1 - \theta) \Delta^{v_2} C \subseteq \Delta^{\theta v_1 + (1 - \theta)v_2} C.$$

Therefore, for each fixed $u \in \mathbb{R}^d$, $s_i \in \Delta^{v_i} C$, $i = 1, 2$, we have

$$\begin{aligned} \text{dist}(u, \Delta^{\theta v_1 + (1 - \theta)v_2} C) &\leq \text{dist}(u, \theta \Delta^{v_1} C + (1 - \theta) \Delta^{v_2} C) \leq \|u - (\theta s_1 + (1 - \theta)s_2)\|_2 \\ &\leq \theta \|u - s_1\|_2 + (1 - \theta) \|u - s_2\|_2 \end{aligned}$$

Taking the infimum over s_1 and s_2 , it follows that for each fixed u , the function $v \rightarrow \text{dist}(u, \Delta^v C)$ is convex. The square of a non-negative convex function is still convex, and since choosing $u = g$ a Gaussian and taking the expectation also does not destroy the convexity, and we can conclude that ϕ is convex.

It remains to prove strict convexity. Since convexity already has been established, it suffices to prove that there does not exist $v_1 \neq v_2$ and $\theta \in (0, 1)$ with $\phi(\theta v_1 + (1 - \theta)v_2) = \theta \phi(v_1) + (1 - \theta) \phi(v_2)$. Towards a contradiction, assume that this is not true. Then

$$\mathbb{E}(\text{dist}^2(g, \Delta^{\theta v_1 + (1 - \theta)v_2} C)) = \mathbb{E}(\theta \text{dist}^2(g, \Delta^{v_1} C) + (1 - \theta) \text{dist}^2(g, \Delta^{v_2} C)).$$

According to what was just proven, the expression over which we are taking the expectation on the right hand side is not smaller than the expression on the left hand side. Hence, in order for equality to be true, the two expressions have to be equal almost surely. But there exists some $u \in \Delta^{\theta v_1 + (1 - \theta)v_2} C$ which does not lie in $\Delta^{v_1} C$ (since $v_1 \neq \theta v_1 + (1 - \theta)v_2$). For this vector, strict equality must hold, since $\text{dist}^2(u, \Delta^{\theta v_1 + (1 - \theta)v_2} C) = 0 < \text{dist}^2(u, \Delta^{v_1} C)$. Now, since the distance from a vector u to a convex set is continuously dependent on u , the strict inequality even holds in some ϵ -ball around u , which has positive Gaussian measure. The two expressions are hence not almost surely equal. This is a contradiction, and hence ϕ is strictly convex.

Continuity: We have, for $u \in \mathbb{R}^d$ fixed and each $p \in C$

$$\text{dist}(u, \Delta^v C) \leq \|u - \Delta^v p\|_2 \leq \|u - \Delta^{\hat{v}} p\|_2 + \sum_{p \in C} \|\Delta^{\hat{v}} p - \Delta^v p\|_2$$

$$\begin{aligned}
|\text{dist}(u, \Delta^v C) - \text{dist}(u, \Delta^{\hat{v}} C)| &\leq \sup_{p \in C} \|(\Delta^v - \Delta^{\hat{v}})p\|_2 \leq \sup_{p \in C} \|\Delta^{v-\hat{v}} p\|_2 \\
&= \sup_{p \in C} \sqrt{\sum_{i=1}^d (v(i) - \hat{v}(i))^2 p(i)^2} \leq \|v - \hat{v}\|_2.
\end{aligned}$$

Denote $\text{dist}(g, \Delta^v C) = r_v$ and $\text{dist}(g, \Delta^{\hat{v}} C) = r_{\hat{v}}$, respectively. The inequality just proved then reads $|r_v - r_{\hat{v}}| \leq \|v - \hat{v}\|_2$ and we may estimate

$$\begin{aligned}
|\phi(v) - \phi(\hat{v})| &= |\mathbb{E}((r_v - r_{\hat{v}})(r_v + r_{\hat{v}}))| \leq \sqrt{\mathbb{E}((r_v - r_{\hat{v}})^2)} \sqrt{\mathbb{E}((r_v + r_{\hat{v}})^2)} \\
&\leq \|v - \hat{v}\|_2 \sqrt{2\mathbb{E}(r_v^2 + r_{\hat{v}}^2)} = \|v - \hat{v}\|_2 \sqrt{2(\phi(v) + \phi(\hat{v}))}.
\end{aligned}$$

ϕ is furthermore locally bounded, since $\phi(v) = \mathbb{E}(\text{dist}^2(g, \Delta^v C)) \leq 2\mathbb{E}(\|g\|^2 + \sup_{s \in C} \|D^v s\|_2^2) = 2 + 2\|v\|_2^2$. Hence, ϕ is continuous.

Coercivity: We have to prove that $\|v\|_2 \rightarrow \infty$ implies $\phi(v) \rightarrow \infty$. Let us first note that due to our global assumption $\alpha_i > 0$ for each i , there exists for every $i = 1, \dots, k$ an index $j_i \in S_i$ with $j_i \in \text{supp } x_0$. For each such index and each $p \in C$, we have due to $p|_{\text{supp } x_0} = \text{sgn } x_0$

$$\inf_{p \in C} \|\Delta^v p\|_2^2 \geq \inf_{p \in C} \sum_{i=1}^k v_i^2 p(j_i)^2 = \sum_{i=1}^k v_i^2 = \|v\|_2^2.$$

That implies that for any given $K > 0$, there exists an $R > 0$ so that $\|v\|_2 \geq R \Rightarrow \inf_{p \in C} \|\Delta^v p\|_2^2 \geq K$. Hence, if $\|v\| \geq R$ and $\|u\| \leq r \leq K$, we have

$$\begin{aligned}
\text{dist}^2(u, \Delta^v C) &\geq \text{pos}\left(\inf_{p \in C} \|\Delta^v p\|_2 - \|u\|_2\right)^2 \\
&\geq (K - r)^2.
\end{aligned}$$

This implies that if $\|v\|_2 \geq R$,

$$\begin{aligned}
\mathbb{E}(\text{dist}^2(g, \Delta^v C))^2 &\geq \mathbb{E}(1_{\{\|g\| \leq r\}} \text{dist}^2(g, \Delta^v C))^2 \\
&\geq \mathbb{P}(\|g\| \leq r) (K - r)^2.
\end{aligned}$$

Since $\mathbb{P}(\|g\| \leq r) > 0$, this proves the claim.

Now it remains to prove that the minimum is not attained in $w = 0$. For this, consider choosing all weights equal to $\lambda > 0$, i.e $\omega = \lambda(1, \dots, 1)$. Lemma 7 applied for $\omega_i = \lambda$ for all i yields

$$\phi(\lambda(1, \dots, 1)) = \sigma + \lambda^2 \sigma + (1 - \sigma) \varphi(\lambda).$$

We may estimate $\varphi(\lambda)$ by

$$\varphi(\lambda) = \sqrt{\frac{2}{\pi}} \int_{\lambda}^{\infty} (x - \lambda)^2 \exp(-x^2/2) dx \leq \sqrt{\frac{2}{\pi}} \int_0^{\infty} (x^2 - 2x\lambda + \lambda^2) \exp(-x^2/2) dx = \left(1 + \lambda^2 - \lambda \sqrt{\frac{2}{\pi}}\right).$$

Therefore

$$\phi(\lambda(1, \dots, 1)) = \sigma + \lambda^2 \sigma + (1 - \sigma) \varphi(\lambda) \leq \sigma + \lambda^2 \sigma + (1 - \sigma) \left(1 + \lambda^2 - \lambda \sqrt{\frac{2}{\pi}}\right) < 1 = \phi(0)$$

for small values of λ , since $\sigma < 1$ by assumption. □

Having established an abstract result on the existence of optimal weights ω , we now take a step back to see that the proof of said theorem actually provides a way to concretely calculate them – we only need to minimize the function $\phi(v) = \mathbb{E}(\text{dist}^2(g, \Delta^v C))$. Writing $v = \tau\omega$ as in the proof, we obtain, with the help of Lemma 7

$$\phi(v) = \sigma + \sum_{i=1}^k \rho_i (\alpha_i v_i^2 + (1 - \alpha_i)\varphi(v_i)).$$

Since ϕ has the structure of a sum of convex functions, it can further be minimized by considering one variable at a time. One could argue that this would have been a much simpler way of proving the assertion of Theorem 10, but the approach used in that proof has more potential to be applied to more general problems, and hence has an interest of its own.

Since the convex functions are even differentiable, it is not hard to write down a formula for calculating $\omega_i = v_i$ analytically. As a matter of fact, this has already been carried out in [1, Prop 4.5]. There, the formula was used in the process of calculating the statistical dimension of $\mathcal{D}(\|\cdot\|_1, x_0)$. In particular, it was not recognized as an optimal choice of a weight ω_i . For completeness, we include the statement.

Corollary 11. *The optimal weights ω_i^* described in Theorem 10 are given through the equations*

$$\alpha_i \omega_i^* = (1 - \alpha_i) \sqrt{\frac{2}{\pi}} \int_{\omega_i^*}^{\infty} (x - \omega_i) \exp(-x^2/2) dx, \quad i = 1 \dots k.$$

In particular, ω^ is independent of ρ .*

Finally, we can extract one more interesting corollary, which in fact is a counterpart of Theorem 2 in this setting. Using the fact that $\sigma = \sum_{i=1}^k \alpha_i \rho_i$, one can see that

$$\phi(v) = \sum_{i=1}^k \rho_i (\alpha_i (1 + v_i^2) + (1 - \alpha_i)\varphi(v_i)).$$

Noticing that the term in the sum is actually the function we need to minimize in order to find $\mu_{\alpha_i|S_i|,|S_i|}$, we immediately arrive at the following result, which also summarizes the entirety of this paper.

Theorem 12. *Let $T \subseteq [1, \dots, d]$ with $|T| = s < d$, and a partition $\mathcal{S} = (S_i)_{i=1}^k$ of $[1, \dots, d]$ be given. Then there exist weights ω^* which minimize $\tilde{m}_{\mathcal{S},T,w}$ and which are unique up to multiplication with a positive scalar. Furthermore, we have*

$$\tilde{m}_{\mathcal{S},T} = \sum_{i=1}^k \rho_i \mu_{\alpha_i|S_i|,|S_i|},$$

and the same also for the lower bounds $\tilde{n}_{\mathcal{S},T} = \tilde{m}_{\mathcal{S},T} - 2/\sqrt{d}$.

4 Numerical experiments

In this section, we present the results of some numerical experiments investigating the practical performance of the weighting strategy described above. Before presenting the set up of the experiment, let us note that it is numerically relatively easy to calculate the optimal weights described in Corollary 11; we simply have to solve the k independent, one-dimensional equations

$$\alpha_i \omega_i^* = (1 - \alpha_i) \sqrt{\frac{2}{\pi}} \int_{\omega_i^*}^{\infty} (x - \omega_i) \exp(-x^2/2) dx, \quad i = 1 \dots k$$

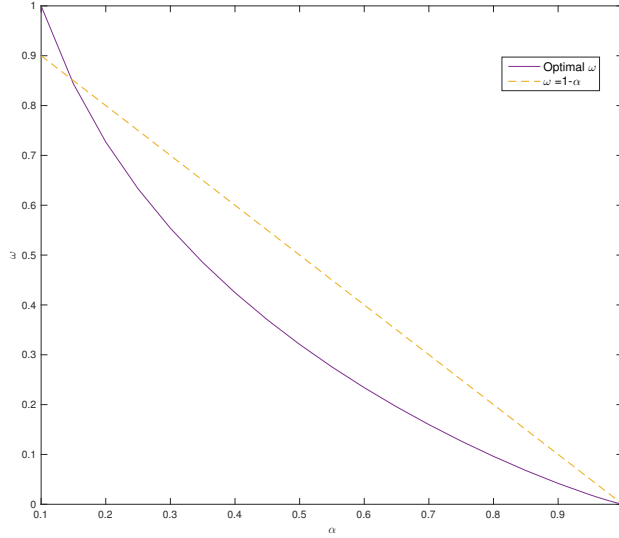


Figure 1: Plot of the optimal weight ω_1 (ω_2 is always equal to 1) for $\rho = (0.1, 0.9)$ for different α_1 compared with $1 - \alpha_1$.

for ω_i . For this, we used the MATLAB routine `fzero`. To be able to compare these weights to other strategies, we furthermore normalized them so that $\|\omega\|_\infty = 1$. In a similar manner, one can explicitly write down the optimality condition for $J'_{\sigma, \alpha, \rho, \omega}(\tau) = 0$ to calculate the thresholds induced by given weights ω .

In the first set of experiments, we consider the case of two sets S_1, S_2 . We fix the ambient dimension to $d = 100$ and choose random 10-sparse signals - the support of x_0 is drawn at random, and the values of x_0 on the support are drawn from the standard multivariate normal distribution. Then we draw 3, 5 or 7, respectively, indices in the support together with 7, 5 or 3, respectively, indices outside the support to form a group S_1 . The rest of the indices is then called S_2 . Hence, in each experiment, $\rho_1 = \sigma = 0.1$ and $\alpha_1 = 0.3, 0.5$ and 0.7 , respectively.

We consider 4 strategies:

1. $\omega = (1, 1)$, which corresponds to standard ℓ_1 -minimization,
2. the extreme strategy $\omega = (0, 1)$,
3. $\omega = (1 - \alpha(1), 1)$, i.e. the one from [7],
4. the one with weights calculated as proposed in this work.

The resulting weights for strategy (4) are depicted in Figure 1 for different values of α . The theoretical thresholds $\inf_{\tau > 0} J_{\sigma, \rho, \alpha, \omega}(\tau)$ were also calculated and are depicted in Figure 2. As can be seen, the statistical dimensions for the our choice are in all cases lower compared to the other approaches. The strategy from [7] is however very close to the one described in this paper.

To test the performance in practice, we for each $m \in [1, \dots, 35]$ draw a matrix $A \in \mathbb{R}^{m, d}$ according to the Gaussian distribution and solve the reweighted ℓ_1 -minimization problem (P_1) with help of the matlab package `cvx` [5]. For each m , 1000 experiments are performed, and a success is declared if the solution of the minimization problem differs no more than 0.001 in ℓ_2 -norm from x_0 . The results can be found in Figure 3.

The figures show that reweighted ℓ_1 -minimization performs significantly better using the weight chosen as in Theorem 10 compared to standard Basis Pursuit ($\omega = 1$) for all α tested. The same is true for the

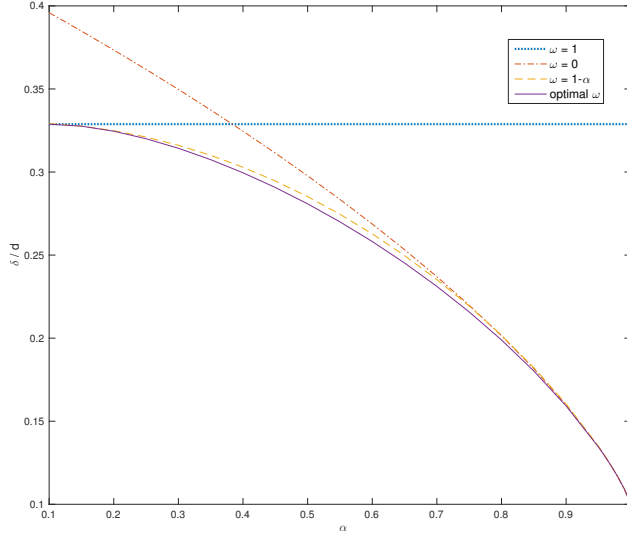


Figure 2: The threshold amount of Gaussian measurements needed to secure recovery with weighted ℓ_1 -minimization for different weighting strategies and $\rho = (0.1, 0.9)$, depending on α_1 .

comparison to the strategy $\omega = 0$ in the cases $\alpha = .3$ and $\alpha = .5$, but not in the case $\alpha_1 = 0.7$. This was expected, since experiments performed by the authors of [3] indicated that the choice of ω gets more important as α gets lower. The difference between our strategy and the one proposed in [7] is also present, however not significant.

In order to see that also the theory for $k > 2$ is practically relevant, we make another experiment and consider the case when $\alpha = (4/5, 3/10, 2/15, 1/70)$ and $\rho = (.05, .1, .15, .7)$. A is chosen as above for different values of m . We consider five different strategies

1. Set $\omega = 1$ (i.e. perform standard ℓ_1 -minimization).
2. Consider the union $S_1 \cup S_2 \cup S_3$ as one region with $\rho = .3$ and $\alpha = 9/30$, and choose the two weights as $\omega = (1 - \alpha, 1)$.
3. Consider the union $S_1 \cup S_2 \cup S_3$ as one region with $\rho = .3$ and $\alpha = 9/30$, and choose the two weights as proposed in this work.
4. Choose four weights, one for each S_i , with $\omega_4 = 1$ and $\omega_i = 1 - \alpha_i$ for $i = 1, \dots, 3$.
5. Choose four weights as proposed in this work.

Note that although strategy (4) bare resemblances to the one proposed in [7], there are no theoretical results which motivate why it should be used. In particular, it was not proposed by the authors of that paper, since they only considered the case $k = 2$. It should only be seen as a heuristic choice for comparison purposes. The results are depicted in Figure 4. We see that the optimal strategy proposed in this work involving four sets S_i is perform significantly better than all of the other strategies.

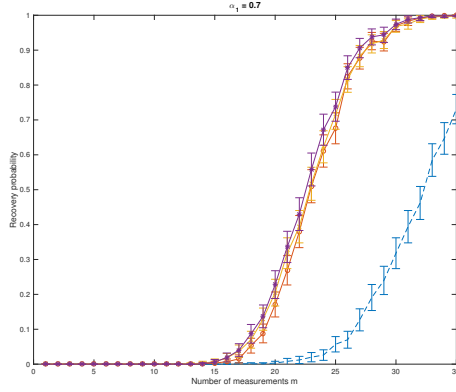
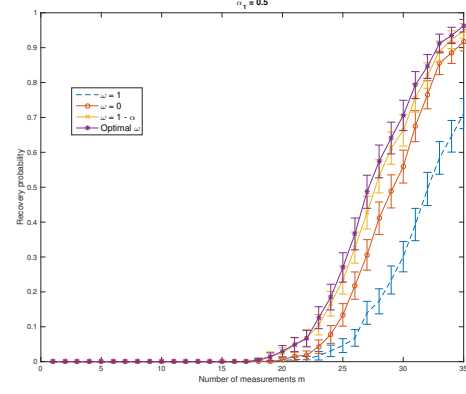
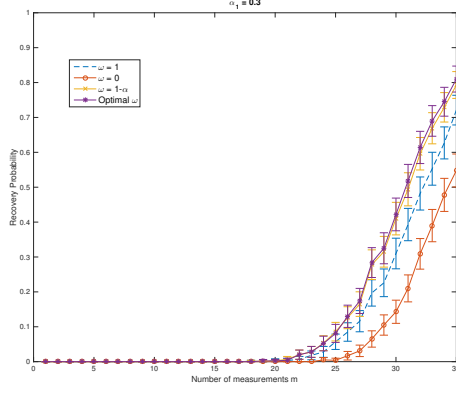


Figure 3: Numerical experiments concerning the effect of optimally choosing ω . In each figure, $d = 100$, $\sigma = \rho_1 = 0.1$. α_1 is different in each figure – 0.3 in the upper left figure, 0.5 in the upper right figure and 0.7 in the lower figure, respectively. The calculated optimal weight ω_1 (ω_2 is always equal to 1) in the respective experiments are equal to 0.5539 ($\alpha_1 = 0.3$), 0.3208 ($\alpha_1 = 0.5$) and 0.1599 ($\alpha_1 = 0.7$). The error bars are corresponding to 3 standard deviations.

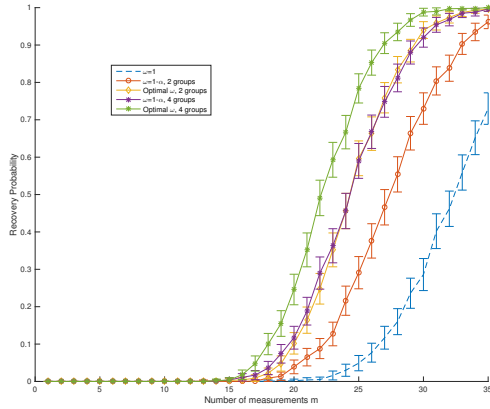


Figure 4: Numerical experiment concerning the effect of using a finer partition S_1, S_2, S_3, S_4 instead of forming two sets $S_1 \cup S_2 \cup S_3$ and S_4 . The sparsity parameter is $\sigma = 0.1$, $\rho = (0.05, 0.1, 0.15, 0.7)$ and the ambient dimension is $d = 100$. The optimal weights were in the first case calculated to $(0.0884, 0.3742, 0.5617, 1)$ and in the second case $(.3742, 1)$. The error bars are corresponding to 3 standard deviations.

Acknowledgment

The author wishes to thank Prof. Dr. G. Kutyniok for carefully proofreading of the paper as well as for fruitful discussions. He also wants to thank the anonymous reviewers for many useful comments. The author also acknowledges support from the Deutscher Akademischer Austauschdienst, DAAD.

References

- [1] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp. Living on the edge: phase transitions in convex programs with random data. *Information and Inference*, 3:224 – 294, 2014.
- [2] E. J. Candès, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. on Pure and Appl. Math.*, 59(8):1207–1223, 2006.
- [3] M. Friedlander, H. Mansour, R. Saab, and O. Yilmaz. Recovering compressively sampled signals using partial support information. *IEEE Trans. Inf. Theory*, 58(2):1122–1134, 2012.
- [4] Y. Gordon. On Milman’s inequality and random subspaces which escape through a mesh in \mathbb{R}^n . In J. Lindenstrauss and V. Milman, editors, *Geometric Aspects of Functional Analysis*, volume 1317 of *Lecture Notes in Mathematics*, pages 84–106. Springer Berlin Heidelberg, 1988.
- [5] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. <http://cvxr.com/cvx>, Mar. 2014.
- [6] M. A. Khajehnejad, W. Xu, A. S. Avestimehr, and B. Hassibi. Weighted ℓ_1 minimization for sparse signals with prior information. In *IEEE Inter. Symp. Inf. Theory, 2009.*, pages 483–487, 2009.
- [7] H. Mansour and R. Saab. Recovery analysis for weighted ℓ_1 -minimization using a null space property. *Appl. Comp. Harm. Anal.*, DOI: 10.1016/j.acha.2015.10.005, 2015
- [8] S. Oymak, M. Khajehnejad, and B. Hassibi. Recovery threshold for optimal weight ℓ_1 -minimization. In *IEEE Inter. Symp. Inf. Theory, 2012.*, pages 2032–2036, 2012.
- [9] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.
- [10] N. Vaswani and W. Lu. Modified-CS: Modifying compressive sensing for problems with partially known support. In *IEEE Inter. Symp. Inf. Theory, 2009*, pages 488–492, 2009.
- [11] W. Xu. *Compressive Sensing for Sparse Approximations: Constructions, Algorithms and Analysis*. PhD thesis, California Institute of Technology, 2010. Available online at thesis.library.caltech.edu/5329/1/Thesis.pdf.